

# An Estimation Method Using Periodic Inspection of Indicators

Zheng Wang

**Abstract**—This paper proposes a new approach for estimating the failure time distribution using the indicator data. The indicators, which are checked by periodic inspection of a standby redundant system, only convey whether at least one failure occurs per interval. The estimation procedure first obtains the estimation of the forward recurrence time using the indicator data. Then the mean is estimated based on its relationship with the forward recurrence time. And the estimation of the sampled Cdf is thus derived based on its relationship with the forward recurrence time and the mean. Finally, the Cdf function is estimated using interpolation method. The simulation results showed that the estimation method performed well for the four Weibull distributions.

**Index Terms**—Renewal process, Indicator, Forward recurrence time.

## NOTATION

$X_i$	inter-event times, $i = 1, 2, \dots$ ; they are i.i.d. random variables
$F(*)$	Cdf of $X_i$
$W$	time from the origin to the next event (forward recurrence time)
$g(*), G(*), \bar{G}(*)$	pdf, Cdf, Sf of $W$
$\mu$	$E\{X\}$ , mean failure time
$A_i$	number of events in the interval $((i-1)t, it]$ , $i = 1, 2, \dots, v$
$v$	number of observation intervals
$a_i$	observed value of $A_i$
$\mathbb{A}_i$	indicator function of $A_i$ , $\mathbb{A}_i \equiv 1$ , if $A_i = 0$ ; $\mathbb{A}_i \equiv 0$ , otherwise
$\mathcal{D}_i$	observed value of $\mathbb{A}_i$
$T$	observation period $(0, vt]$
$t$	fixed interval for observations

## I. INTRODUCTION

Periodic inspection is one of the most common testing or maintenance activities, which are basically designed for estimating or improving the reliability of a system [1],[2],[3],[4]. With the purpose of detecting failures, periodic inspection is typically undertaken regularly at a constant rate while the system is continuously operated. That means the system is not interrupted by the failures, thereby the inter-failure times are not impacted by the inspection activities. Compared with continuous inspection, the failures can be detected not immediately but only at the next inspection once they occur. In other words, there is some delay between the real occurrence of a failure and its detection.

The uninterruptibility of the system is ensured at least in two cases:

- The system has enough redundancy so that upon failure of one component in function, the backup is automatically activated as a replacement in a negligible amount of time. Such redundant systems were well studied and documented [5],[6],[7].
- The system has components with hidden or soft failures. While so-called hard failures make the system stop functioning as soon as they occur, the system can continue to operate when one or more soft failures take place. Those components with soft failures include protective components, standby components, and secondary components, which do not carry out the main functions of a system [8],[9].

Despite of the uninterruptibility of the system, the consumption of redundant components or the presence of components with soft failures, if accumulated over time and not detected nor addressed, can still eventually reduce the performance or even break the operation of the system. For example, the system may fail if the redundant components are used up, or the number of components with soft failures amount to a certain threshold. So periodic inspection are performed to detect and/or rectify the failures and to ensure that the system is safe and reliable. Typical rectification efforts taken by periodic inspection includes remedying the loss of redundancy and fixing the components with failures.

A strong assumption on the data provided by periodic inspection would be the count data, number of failures per interval. Dattero and White [10] proposed an estimation approach based on periodic inspection of the count data. However, the count data is not always known or provided in periodic inspection. In a wide diverse of systems, there is no failure counter dedicated for the inspection. Some inspections, especially those performed passively or by outsiders, are much like black-box testing where only limited or even minimum information about failures can be retrieved. This paper considers an indicator on which the estimation of the failure-time distribution is based. The indicator only tells whether one or more failures occur per interval rather than the exact number of failures. The data provided by indicator may be most readily available for the monotonically evolved parameters of a system.

For example, since the accumulated failure repair time of a system is non-decreasing, a change in the accumulated failure repair time from the previous inspection would serve as the indicator that there is one or more failures in the interval (but

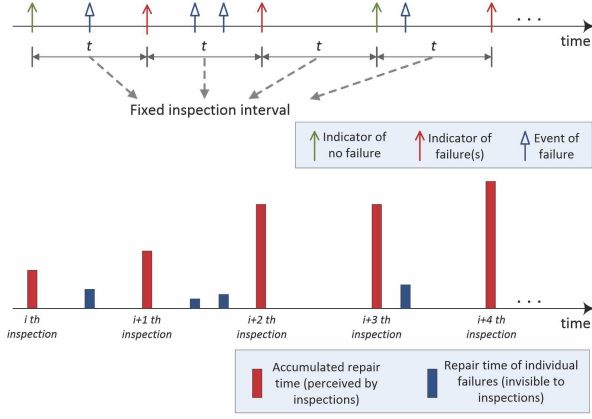


Fig. 1. Periodic inspection and an example of accumulated failure repair time as the indicator

the number of failures are never known because the repair time of an individual failure may vary). The example is illustrated in Fig.1, where the  $i+1$  th,  $i+2$  th, and  $i+4$  th inspections observe an increase in the accumulated failure time caused by one, two, and one failure(s) respectively. By comparison, the  $i+3$  th inspection finds no change in the accumulated failure time in Fig.1, indicating no failure occurs between the  $i+2$  th inspection and the  $i+3$  th inspection.

The incapability of obtaining the number of failures may result from the privacy concerns or the lack of continuous inspections. Some systems ensure the privacy of their operations by leaking a minimum information where some sensitive data such as the number of failures are kept private and unavailable for outsiders. Some systems cannot afford the cost of continuous inspections so that the failures are not one-by-one thoroughly identified and counted.

This paper provides an estimation method using periodic inspection of indicators. Based on the indicator function of failures observed at fixed time intervals, the failure-time distribution is estimated. First, the estimation process is formally developed. Then the estimation procedure is stated. Finally, the performance of the estimation method is assessed using simulated data for various Weibull distributions.

## II. PROBLEM DEFINITION

Assumptions:

- There is a stationary renewal process.
- The inter-event times are i.i.d. random variables.
- Data are collected on a standby redundant system over a fixed number of collection intervals.
- The goal is to find an estimator for the Cdf of inter-event times.

The equation for the stationary renewal process to be estimated is [11]

$$F(x) = 1 - \mu g(x) \quad (1)$$

## III. ESTIMATION PROCEDURE

Given (1), a reasonable estimator for  $F(x)$  would seem to be:

$$\tilde{F}(x) = 1 - \hat{\mu} \hat{g}(x) \quad (2)$$

In developing  $\tilde{F}$ , difficulties arise, however, because only event occurrence for each interval is known; neither the actual inter-event times  $X_1, X_2, \dots$  nor the event counts for each interval are known.

### A. Estimator for $\mu, g(*), G(*), \bar{G}(*)$ at $kt$

If the event counts for each interval are known, the mean inter-event time,  $\mu$ , can be straightforwardly estimated by dividing the total period,  $T$ , by the sum of the observed event counts. However, independently estimating  $\mu$  is more much difficult, since the event counts for each interval are not known. So estimating  $\mu$  may not be fully decoupled from estimating  $g(x)$ . Instead this paper estimates  $\mu$  and  $g(x)$  simultaneously. The associated Sf is assessed at  $kt$  as these are the only points where the data provide useful information. The Sf,  $\bar{G}(*)$  is first estimated using

$$\Pi(kt) = \sum_{i=1}^{v-k+1} B_k(i)/(v-k+1) \quad k = 1, 2, \dots, K \quad (3)$$

Notation:

$$\Pi(0) = 1$$

$$B_k(i)$$

$$\Pi(kt)$$

an indicator;  $B_k(i) \equiv \prod_{j=i}^{i+k-1} A_j$

proportion of time no events are reported over  $k$  consecutive intervals each of length  $t$

$$\hat{p}(kt)$$

$$b_k(i)$$

an observation of  $\Pi(kt)$

an observation of  $B_k(i)$

Note that a)  $B_1(i) = 1$  if  $A_i = 0$ , b) overlapping intervals are used (for  $k = 2, 3, \dots, K$ ) in (3), and c)  $K$  is selected such that  $K \leq v$ .

The estimator  $\Pi(kt)$  has at least three desirable properties:

- It is an unbiased estimator (even though the  $B_k(i)$  terms are not  $s$ -independent, the mean value of the sum is still equal to the sum of the mean values of each  $B_k(i)$  term and  $E\{B_k(i)\} = \Pr\{W > kt\}$  for all  $i$ ..)
- It has a recursive method of calculation (since  $B_k(i) = B_{k-1}(i)B_{k-1}(i+1)$  for  $k = 2, 3, \dots$ ).
- It is monotonically nonincreasing in  $k$ .

For formal proofs of these properties and some more esoteric results, see [12]. The  $g(kt)$  is estimated using the centered difference equation to estimate a derivative for  $k = 1, 2, \dots, K-1$ :

$$\hat{g}(kt) = [\hat{p}((k-1)t) - \hat{p}((k+1)t)]/2t \quad kt > 0$$

$$\hat{g}(0) = 1/\hat{\mu} \quad (4)$$

$$\lim_{x \rightarrow \infty} \hat{g}(x) = 0 \quad (5)$$

since these properties hold for  $g(x)$ . To ensure:  $\lim_{x \rightarrow \infty} \hat{g}(x) = 0$ , the  $K$  is usually chosen in such a way that  $\hat{p}((K-2)t) = \hat{p}((K-1)t) = \hat{p}(Kt) = 0$ . Using  $\hat{g}(kt)$  from (4),  $\tilde{F}(kt)$  from (2) is not necessarily monotonic. Therefore,  $F(kt)$  is estimated using

$$\tilde{F}(kt) = \text{MAX}\{\tilde{F}(kt), \tilde{F}((k-1)t)\} \quad (6)$$

$$\tilde{F}(0) \equiv 0 \quad (7)$$

Using this procedure ensures that the sequence  $\tilde{F}(t), \tilde{F}(2t), \dots, \tilde{F}((K-1)t)$  is monotonically nondecreasing and that  $0 \leq \tilde{F}(kt) \leq 1, k = 1, 2, \dots, K-1$ .

According to (4), we need to derive  $\hat{g}(0)$  in order to have  $\hat{\mu}$ . We have

$$\int_0^\infty g(x) = 1 \quad (8)$$

which can be written as

$$\sum_{k=1}^\infty \int_{(k-1)t}^{kt} g(x) = 1 \quad (9)$$

The integrals in (9) can be estimated using  $\hat{g}(kt)$

$$\sum_{k=1}^\infty [\hat{g}(kt) + \hat{g}((k-1)t)]t/2 = 1 \quad (10)$$

Since  $\hat{g}(kt) = 0$ , for  $k = K-1, K, \dots$ , we have

$$\sum_{k=1}^{K-1} (\hat{g}(kt) + \hat{g}((k-1)t))t/2 = 1 \quad (11)$$

So  $\hat{g}(0)$  can be derived from the following

$$\hat{g}(0)t/2 + \sum_{k=1}^{K-2} \hat{g}(kt)t = 1 \quad (12)$$

Then  $\tilde{F}(kt)$  can be estimated per (2).

### B. Estimator for $g(*)$ , $G(*)$ , $\overline{G}(*)$ at $x$

The final step in the procedure is the estimation of this function for  $x > 0$ , but  $x \neq t, \dots, (K-1)t$ . The recommended procedure is linear interpolation so that for  $(k-1)t < x < kt$ :

$$\tilde{F}(x) = \tilde{F}((k-1)t) + (x - (k-1)t)[\tilde{F}(kt) - \tilde{F}((k-1)t)]/t \quad (13)$$

Linear interpolation has the advantages of simplicity and ease of use. However, any other interpolation method can be used, providing it gives a monotonic estimate of  $F(x)$ .

### C. Algorithm

For the sake of clarity, we give an algorithm that pulls together the segments discussed above and given in (2) - (12).

#### 1. CALCULATE

- $v = T/t$

#### 2. DEFINE

- $\hat{p}(0) = 1$
- $\tilde{F}(0) = 0$

#### 3. FOR $k = 1, 2, \dots, v$

- For  $i = 1, 2, \dots, (v - k + 1)$   
If  $k = 1$  SET  $b_1(i) = 1$  IF  $a_i = 0$  OR SET  $b_1(i) = 0$  IF  $a_i > 0$   
If  $k > 1$  CALCULATE  $b_k(i) = b_{k-1}(i)b_{k-1}(i+1)$
- CALCULATE  $\hat{p}(kt) = \sum_{i=1}^{v-k+1} b_k(i)/(v - k + 1)$

#### 4. DETERMINE

- $K = \arg \min_{x=1,2,\dots,v-1} \hat{p}((x-2)t) - \hat{p}((x-1)t) = \hat{p}(xt) = 0$
- 5. FOR  $k = 1, 2, \dots, K-1$ 
  - CALCULATE  $\hat{g}(kt) = [\hat{p}((k-1)t) - \hat{p}((k+1)t)]/2t$
  - CALCULATE  $\hat{\mu} = t/[2(1 - \sum_{k=1}^{K-2} \hat{g}(kt)t)]$
  - CALCULATE  $\tilde{F}(kt) = 1 - \hat{\mu}\hat{g}(kt)$
  - DETERMINE  $\tilde{F}(kt) = \text{MAX}\{\tilde{F}(kt), \tilde{F}((k-1)t)\}$
- 6. FOR  $x > 0$ , but  $x \neq t, 2t, \dots, (K-1)t$ 
  - LINEARLY INTERPOLATE FOR  $(k-1)t < x < kt$   
 $\tilde{F}(x) = \tilde{F}((k-1)t) + (x - (k-1)t)[\tilde{F}(kt) - \tilde{F}((k-1)t)]/t$

## IV. PERFORMANCE EVALUATION

While the most desirable evaluation is based on the closed form analytic results of the estimation method, the complexity of the joint distributions involved in the proposed estimation procedure makes it hard to derive the closed form expression. For example, few closed form analytic results are readily available for the joint distributions of the indicators. The computation of  $\hat{g}(kt)$  and  $\tilde{F}(x)$ , especially when compound with the indicators, adds to the complexity.

Since the statistical complexity of the proposed estimation procedure results in the hardness of the closed form analytic evaluation, we rely on Monte Carlo simulation for evaluations. The Weibull distribution has a variety of distribution shapes and wide use in reliability studies. So we choose four Weibull distributions in the simulations, among which one special case is the exponential distribution. As illustrated in Table 1, the four Weibull distributions all have their means around 1.

TABLE I  
PARAMETERS OF WEIBULL DISTRIBUTION  
 $F(x) = 1 - \exp[-(x/\alpha)^\beta]$

Distribution	Scale $\alpha$	Shape $\beta$	Mean
1	1.090	5.0	1.001
2	1.009	3.5	0.908
3(Exponential)	1.000	1.0	1.000
4	0.878	0.8	0.995

For each of the four distributions, we generated 1000 independent runs of event epochs. For each run, estimates were made for periods  $T$  of 50, 100, 500, 1000. Each period  $T$  was sectioned into periodic intervals of fixed length  $t$  of 0.1, 0.2, 0.5, 1 and the indication data  $t$  for each interval were recorded. We obtained the Cdf estimation for each run, for each period of length  $T$ , and for each interval of length  $t$ .

The performance metric used was the maximum absolute distance between the estimated Cdf and actual Cdf,  $|\hat{F}(x) - F(x)|$ . Thus, in cases where the measure is small, the estimation procedure closely approximates the actual Cdf. As the intermediate result of the final estimation, the estimated mean may also serve as a parameter interested and desired in many cases. So we also used the absolute distance between the estimated mean and actual mean,  $|\hat{\mu} - \mu|$ , as the other performance metric. When this measure is small, the estimation procedure successfully approximates the actual mean.

Table 2 summarizes the means of the maximum absolute Cdf differences for the various combination of run length,

$T$ , and  $t$  for each of the distributions. In general, we can see that the estimation method shows good performance for all combinations of setting. Distribution (1) and Distribution (2) have similar best performance (Distribution (1) is better for some settings and Distribution (2) is better for the other settings). The estimation method performed worst for Distribution (4). The explanation may be that the estimation method is relatively less favorable for distributions that have high probability mass concentrated in a small interval. This comparative performance fall is caused by two operations in the estimation procedure: 1) the derivation of  $\hat{g}(0)$  (thus  $\mu$ ) based on  $\hat{g}(kt)$  assuming that the integral of the probability distribution of the forward recurrence time can be approximated using  $\hat{g}(kt)$ ; 2) the linear interpolation which is a sub-optimal interpolation method for some distributions.

TABLE II  
MEANS OF MAXIMUM ABSOLUTE CDF DIFFERENCE

$T$	$t$	Distribution			
		(1)	(2)	(3)	(4)
50	0.1	.078	.090	.101	.123
	0.2	.078	.071	.119	.163
	0.5	.181	.142	.200	.269
	1	.331	.315	.321	.373
100	0.1	.055	.056	.081	.110
	0.2	.065	.052	.096	.153
	0.5	.178	.132	.197	.261
	1	.325	.314	.313	.371
500	0.1	.028	.056	.081	.110
	0.2	.053	.036	.091	.153
	0.5	.168	.130	.197	.265
	1	.323	.309	.316	.367
1000	0.1	.022	.019	.050	.095
	0.2	.052	.034	.091	.152
	0.5	.168	.130	.196	.265
	1	.323	.309	.317	.367

Factor Means						
Distribution	Mean	$T$	Mean	$t$	Mean	Grand Mean
(1)	.152	50	.185	0.1	.072	.172
(2)	.137	100	.172	0.2	.091	
(3)	.173	500	.168	0.5	.192	
(4)	.225	1000	.162	1	.331	

In Table 2, we can observe the performance degrade with the increase of the interval  $t$  for most of the results. This seems accord with our intuition because more detail is lost for a longer interval. Another finding is that the increase of the observation period  $T$  contributes to some improvement of estimation. This also matches our intuitive expectations that more samples facilitates a higher accuracy of estimation. However, a smaller value of  $t$  does not always provide a better estimate. For example, Distribution (2) shows a better estimate for  $t = 0.2$  than for  $t = 0.1$  when the observation period  $T$  is small. The reason of this anomaly mainly lies in the errors introduced by estimating the integral of the probability distribution of the forward recurrence time using  $\hat{g}(kt)$  and the sub-optimality of the linear interpolation. The performance of linear interpolation varies for different distributions, so there is some space of improvements if some prior knowledge

about the shape of the underlying distribution is known and better interpolation methods can be found accordingly. Another anomaly is no performance improvement with the grow of  $T$  in Distribution (3). This phenomena also results from estimating the integral of the probability distribution and the linear interpolation.

Table 3 summarizes the absolute mean differences for the various combination of run length,  $T$ , and  $t$  for each of the distributions. In general, the estimation method performs well for the mean estimation. In terms of the average performance, smaller sampling interval  $t$  and longer observation period  $T$  both help to enhance the performance. Like the Cdf estimation, some anomalies can be found because of the errors introduced by estimating the integral of the probability distribution of the forward recurrence time using  $\hat{g}(kt)$ .

TABLE III  
MEANS OF ABSOLUTE MEAN DIFFERENCE

$T$	$t$	Distribution			
		(1)	(2)	(3)	(4)
50	0.1	.005	.013	.066	.131
	0.2	.007	.014	.122	.188
	0.5	.015	.033	.286	.434
	1	.117	.196	.597	.784
100	0.1	.004	.000	.071	.109
	0.2	.006	.002	.094	.188
	0.5	.010	.020	.274	.395
	1	.106	.195	.575	.769
500	0.1	.000	.003	.057	.113
	0.2	.001	.001	.102	.183
	0.5	.004	.018	.281	.404
	1	.102	.185	.583	.762
1000	0.1	.001	.000	.052	.109
	0.2	.001	.003	.102	.188
	0.5	.004	.017	.273	.400
	1	.102	.185	.581	.759

Factor Means						
Distribution	Mean	$T$	Mean	$t$	Mean	Grand Mean
(1)	.030	50	.188	0.1	.046	.178
(2)	.055	100	.176	0.2	.075	
(3)	.257	500	.175	0.5	.180	
(4)	.370	1000	.174	1	.412	

## V. CONCLUSION

This paper proposed a new approach for estimating the failure time distribution using the indicator data obtained by periodic inspections. Simulations showed that the estimation method performed well for the four Weibull distributions. The proposed estimation method can be applied in the system testing or maintenance practices where only the indicator data rather than any more details are available or desired.

## REFERENCES

- [1] S. Taghipour, D. Banjevic, Periodic Inspection Optimization Models for a Repairable System Subject to Hidden Failures, IEEE Transactions on Reliability, No. 60(1), 275-285, 2011.
- [2] G. Levitin, L. Xing, Y. Dai, Heterogeneous Non-Repairable Warm Standby Systems With Periodic Inspections, IEEE Transactions on Reliability, 2015.

- [3] T.S. Bruggemann, J.J. Ford, Guidance of Aircraft in Periodic Inspection Tasks, Australian Control Conference (AUCC'2011), 445- 451, 2011.
- [4] S. Taghipour, M.L. Kassaei, Periodic Inspection Optimization of a k-Out-of-n Load-Sharing System, IEEE Transactions on Reliability, No. 64(3), 1116-1127, 2015.
- [5] R. Subramanian, V. Anantharaman, Reliability Analysis of a Complex Standby Redundant Systems, Reliability Engineering & System Safety, No. 48(1), 5770, 1995.
- [6] H. Yua, C. Chua, E. Chteleta, F. Yalaouia, Reliability optimization of a redundant system with failure dependencies, Reliability Engineering & System Safety, No. 92(12), 1627-1634, 2005.
- [7] S. Mitra, N.R. Saxena, E.J. McCluskey, A Design Diversity Metric and Analysis of Redundant Systems, IEEE Transactions on Computers, No. 51(5), 498-510, 2002.
- [8] S. Taghipour, D. Banjevic, Optimum Inspection Interval for a System under Periodic and Opportunistic Inspections, IIE Transactions, No. 44(11), 932-948, 2012.
- [9] S. Taghipour, D. Banjevic, A.K.S. Jardine, Periodic Inspection Optimization Model for a Complex Repairable System. Reliability Engineering and System Safety, No. 95(9), 944-952, 2010.
- [10] R. Dattero, E.M. White, A New Estimation Approach Based on Periodic inspection, IEEE Transactions on Reliability, No. 38(4), 436-439, 1989.
- [11] D. R. Cox, P. A. Lewis, *The Statistical Analysis of Series of Events*, Methuen, 1966.
- [12] R. Dattero, *Stochastic Models from Event Count Data*, PhD Thesis, Purdue University, 1982.